DOCUMENT RESUME

ED 025 740 AL 001 373

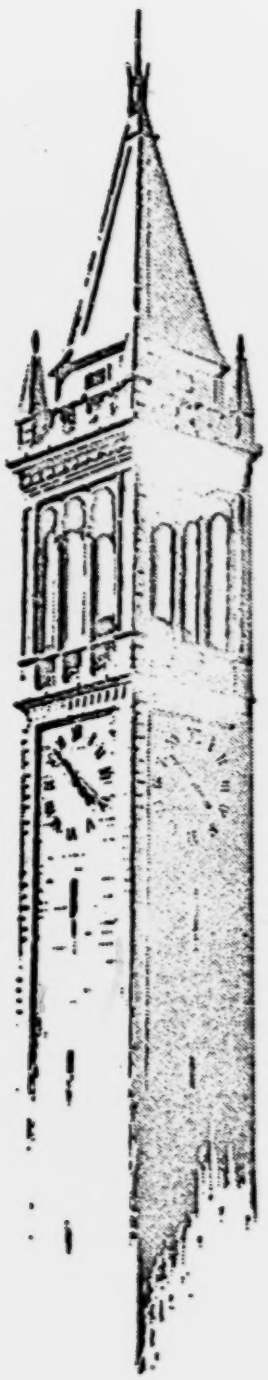
By-Lyovin, Anatole
A Chinese Dialect Dictionary on Computer: Progress Report.
California Univ., Berkeley. Phonology Lab.
Spons Agency-National Science Foundation, Washington, D.C.
Report No-POLA-2-7
Pub Date Jun 68
Note-45p.; Paper in Project on Linguistic Analysis, Reports. Second Series, No. 7.
EDRS Price MF-$0.25 HC-$2.35
Descriptors-Cantonese, *Chinese, Computational Linguistics, Contrastive Linguistics, Diachronic Linguistics, Dialects, Dialect Studies, Dictionaries, Japanese, Korean, *Mandarin Chinese Phonology, *Regional Dialects, Tone Languages
Identifiers-Annamese, Guang Yun, *Hanyu Fangyin Zihui, Ji Yun, Logographs

The use of computers makes possible analysis of the vast amount of data available in recent dialect dictionaries and surveys and in the ancient Chinese rhyme books, such as "Guang yun" and "Ji yun." Comparison of dialects can enable a historical study of Chinese, a major language group outside the Indo-European area, to offer "a more balanced perspective on the nature of sound change in human language." The problems of coding are great, but once the coding system is established, the encoding of materials can be shared by a number of institutions. The coding of the seven Mandarin dialects in the "Hanjy fangyin zihui" is complete, and preliminary tests of the computer program have shown it to be satisfactory. After further testing and refinement, other dialect surveys, rhyme books, and Sino-Korean, Sino-Japanese, and Sino-Annamese can be added to the system. The author gives details of the organization of the data, the coding system, the computer program, and errors in the source data. Appendices give the actual computer code, flow charts for the computer program, and a list of errors in the "Hanjy fangyin Zihui." Correspondence concerning POLA matters should be addressed to William S-Y. Wang, Department of Linguistics, University of California, Berkeley, California 94720. (MK)

for NAR EDITORS

ED025740

PHONOLOGY LABORATORY – DEPARTMENT OF LINGUISTICS
UNIVERSITY OF CALIFORNIA, BERKELEY



Project On Linguistic Analysis

Reports. Second Series, No. 7 : June, 1968

Haruo Aoki. A note on glottalized consonants A1-A13
Kun Chang and Betty Shefts Chang. Vowel harmony in spoken Lhasa Tibetan 1-81
Margaret Lauritsen. A phonetic study of the Danish stød D1-D12
Anatole Lyovin. A Chinese dialect dictionary on computer: Progress report C1-C43
Anatole Lyovin. Notes on the addition of final stops in Maru L1-L22

AL 001 373

48 p

Reproduction in whole or in part is permitted for any purpose of the United States Government.

The Project on Linguistic Analysis is supported in part by the National Science Foundation (Grant GS1430), the Office of Naval Research (Contract N00014-67-A-0114-0005), and the Air Force (Contract F30602-67-C-0347). It is administered through the Phonology Laboratory of the University of California at Berkeley, which has its office in 51 Dwinelle Hall (Telephone: 845-6000, Extension 1507).

Correspondence concerning POLA matters should be addressed to William S-Y. Wang, Department of Linguistics, University of California, Berkeley, California 94720.

A Chinese Dialect Dictionary on Computer: Progress Report*

Anatole Lyovin

University of California, Berkeley

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

*The preparation of this paper was supported by National Science Foundation Grant GS1430.

AL 001 373

Table of Contents

0. Introduction 1

1. The organization of the data in the Hànyǔ fāngyīn zìhuì and our coding scheme 3

2. General notes on the coding format 8

3. The computer program 8

3.0. Purpose 8

3.1. Organization 9

3.2. Method 11

4. Errors in the Zìhuì. 15

Appendix I. Computer code for the Hànyǔ fāngyīn zìhuì 22

Appendix II. David Forthoffer. Flowcharts for the computer program 31

Appendix III. Hsin-i Hsieh. A list of the errata in the Hànyǔ fāngyīn zìhuì 33

References 42

Footnotes 43

0. Introduction.[1]

The project on computerizing Chinese dialect data had its inception in the summer of 1966 at the University of California, Berkeley. At that time, Professor William S-Y. Wang, in his two workpapers,[2] outlined the need for a more viable method of handling the vast amounts of data which are available in the Ancient Chinese rime books, for example, Guǎng yùn and Jí yùn (the latter containing well over 50,000 logographic entries), and in the dialect dictionaries and surveys which have been recently compiled in China. These materials constitute a vast source of data which has to be analyzed in greater detail in order that our reconstruction of the earlier stages of the Chinese language may be more accurate and complete. Furthermore, because until now the most impressive achievements of comparative linguistics have been mainly in the field of Indo-European, a clearer picture of the historical developments in another major language group such as Chinese (which is structurally different from Indo-European) would be crucial for a more balanced perspective on the nature of sound change in human language.

However, the very vastness of the available data on Chinese has until now been more of a hindrance than a help, since the difficulties in tabulating and analyzing this data are almost insurmountable for individual scholars, or even small groups of scholars. The solution to this problem lies in the use of digital computers, which can handle great amounts of data at great speed. Although the initial outlay of time and effort involved is still by no means minimal even for computerizing a single rime or dialect dictionary, it is much less time-consuming than manual tabulation of information. Moreover, the task of coding different source materials can be distributed to several groups or institutions,

which in return for the money and effort invested will eventually be able to share in the information provided by the computer. In this way, we can begin integrating the materials from diverse sources into a computerized pool of information which will yield factual answers quickly and accurately.

During the summer of 1966, two concrete steps were taken to explore the problems involved in the coding of the data. First, Professor Wang devised a coding system for the material in the Hànyǔ fāngyīn zìhuì[3] (hereafter referred to as the Zìhuì) which will be described below in detail. Second, Charles N. Li, a graduate student at the University of California, Berkeley, was assigned to devise a computer code for Chinese logographs.

In October, 1967, work began on the coding of the seven Mandarin dialects in the Zìhuì. At the time of this writing the coding of the above-mentioned portion of the Zìhuì has been completed; the keypunching of the IBM cards is still in progress, but a sufficient sample of the data cards is available to test the computer programs which have already been written.

After the computer programs have been sufficiently tested on the Mandarin data and found compatible with our purposes, the rest of the dialect data in the Zìhuì will be coded and added. In the future, data from other dialect surveys, rime books, Sino-Korean, Sino-Japanese, and Sino-Annamese can also be added without any need for major revision of the computer programs which have been written.

1. The organization of the data in the Hǎnyǔ fāngyīn zìhuì and our coding scheme.

The Zìhuì contains, roughly, 2700 Chinese logographs with their Ancient Chinese classification according to the rime dictionaries Guǎng yùn and Jí yùn, and their phonetic value in the following seventeen Chinese dialects: Pekinese, Jī-nán, Xī-ān, Tài-yuán, Hàn-kǒu, Chéng-dū, Yáng-zhōu (Mandarin dialects); Sū-zhōu, Wēn-zhōu (Wu dialects); Cháng-shā, Shuāng-fēng (Xiāng dialects); Nán-chāng (a Gàn dialect); Méi-xiàn (a Hakka dialect); Guǎng-dōng (a Cantonese dialect); Xià-mén, Cháo-zhōu (Southern Min dialects); and Fù-zhōu (a Northern Min dialect). (The logographs in the Zìhuì are arranged according to the Peking pronunciation by final, initial, and tone.)

This information in the Zìhuì is arranged in, roughly, 2700 columns; each column contains nineteen cells. Cell 1 contains the Chinese logograph, and sometimes the following information:

(1) If the same logograph appears twice in the Zìhuì (i.e. if it has more than one pronunciation according to its meaning or according to a particular disyllabic word in which it appears as an element), the compilers also list the expression in which the said logograph has a particular phonetic value. For example, on page 117 we have 給 which is pronounced ᶜkei in the Peking Mandarin expression 給你 ; on page 64, we encounter the same logograph with the phonetic value ᶜtɕi which it has in the word 供給 (in the same dialect).

(2) Logographs which have the same phonetic value in all the dialects listed, as well as the same classification in the Ancient Chinese rime books (i.e. logographs completely homophonous with the main logograph in

Cell 1), are listed in a footnote.

The Coding of Cell 1:

Although the code for the logographs which was devised by Charles Li[4] was ingenious, it still failed to present a system whereby a Chinese logograph could easily be coded and decoded. Moreover, his code was not very economical, in that it required long strings of code characters which would substantially increase the time consumed in the keypunching of the data. The Chinese telegraphic code[5] was, therefore, adopted for the coding of Cell 1. This latter code, although still not very satisfactory in many respects, was found to be more useful for our purposes than Li's code. In his final report (Li 1967), Li summarizes the problems involved:

'The problems involved in devising a character coding system are of three types. The first type of problems has to do with the establishment of an isomorphic relation between the codes and the characters; the second type is concerned with decoding; the third type is concerned with simplicity. Thus, a perfect coding system will provide a unique code for each character, a simple and easy to learn coding procedure, and a direct, simple decoding procedure that requires no reference to a dictionary. One may choose to solve only one type of problems and ignore the others. For example, the telegraphic code is a system which provides good solutions to problems of the first and second type, but completely ignores the problems of the second type. My coding system aims at a perfect solution to all the problems, but falls short of its aim. The system has been improved in its ability to provide a unique and a shorter code for

each character. But, the coding and, therefore, the decoding procedures as well, since they are merely the inverse of each other, are getting more complicated. It seems that the only perfect solution to the coding problem is to employ mechanical means to perform both the coding and the decoding...'

In the case of the logographs which appear more than once in the Zìhuì (see p. 3), we have added supplementary alphabetic symbols to the telegraphic code to indicate that a particular logograph appears more than once. For example, the telegraphic code for 給 is 4822; its first occurrence (on p. 64) was accordingly coded as 4822A, whereas its second occurrence (on p. 117) was coded as 4822B. (It should be noted that the telegraphic code sometimes also contains an alphabetic code letter; for example, 糗 is 475A. However, since the telegraphic code always contains four code characters, there will be no confusion between, say, 475A, and 0475A, 0475B: it is only the fifth code character that specifies a case of multiple occurrence in the Zìhuì.)

Finally, the homophonous logographs are coded separately within their own matrices or columns. (In other words, they are treated as separate entries.) We could have easily provided more parts for Cell 1, each part containing the telegraphic code for the homophonous logographs. However, we had to foresee the possibility that the logographs in question might not turn out to be perfectly homophonous to each other in the dialects covered by other dictionaries or surveys which will be coded in the future.

Cell 2 and its coding:

Cell 2 contains information on the phonological categories assigned to each particular logograph by the Ancient Chinese rime books (Guǎng yùn

and Jí yùn). In a few cases where a logograph does not appear in these rime books, Cell 2 is blank and is coded as zero.

This cell contains the following parts:

Part 1. 16 shè (攝) or 'rimemes'. The 16 shè are each coded by two letters: the first letter denotes either nèi zhuàn (內轉)(N, O, P) or wài zhuàn (外轉) (W, X); the second letter denotes the ending (O, I, U, M, N, G). (See Appendix I, Part I.a.)

Part 2. kāi-kǒu vs. hé-kǒu (開口 vs.合口). These are coded as KAI and HE, respectively. (See Appendix I, Part I.b.)

Part 3. Four divisions or děng (等). These are coded as 1D, 2D, 3D, 4D. (See Appendix I, Part I.c.)

Part 4. Tone. 平 = 1, 上 = 2, 去 = 3, 入 = 4. (See Appendix I, Part I.d.)

Part 5. Subrimes or yùn (韻). There are about 189 yùn. These are coded numerically according to the shè to which they belong. (See Appendix I, Part I.e.)

Part 6. Initial or niǔ (紐) (40 categories). These are coded by an alphabetical code based on Dǒng Tóng-hé's reconstruction of the Ancient Chinese initials.[6] (See Appendix I, Part I.f.)

Cells 3-19:

These cells contain the information (in the IPA transcription) on the phonetic value of each logograph in the following dialects:

Cell 3 = Peking (北京)

Cell 4 = Jǐ-nán (濟南)

Cell 5 = Xī-ān (西安)

Cell 6 = Tài-yuán (太原)

Cell 7 = Hàn-kǒu (漢口)

Cell 8 = Chéng-dū (成都)

Cell 9 = Yáng-zhōu (揚州)

Cell 10 = Sū-zhōu (蘇州)

Cell 11 = Wēn-zhōu (温州)

Cell 12 = Cháng-shā (長沙)

Cell 13 = Shuāng-fēng (雙峯)

Cell 14 = Nán-chāng (南昌)

Cell 15 = Méi-xiàn (梅縣)

Cell 16 = Guǎng-zhōu (廣州)

Cell 17 = Xià-mén (厦門)

Cell 18 = Cháo-zhōu (潮州)

Cell 19 = Fù-zhōu (福州)

As stated above, we have so far coded only the seven Mandarin dialects, i.e. Cells 3 to 9 inclusive.

In each cell there may be more than one pronunciation of the logograph: sometimes there are several variant readings of a particular logograph in a single dialect. In many cases the difference among the variant pronunciations is that between the colloquial pronunciation vs. the reading or literary pronunciation. In the Zìhuì, the contrast between the colloquial and reading pronunciation is indicated by a single underline and a double underline, respectively.

In our code each syllable is broken down into four parts:

Part 1. Tone. (For the symbols used, see Appendix I, Part II.a.)

Part 2. Initial. (For the symbols used, see Appendix I, Part II.b.)

Part 3. Vowel complex. (For the symbols used, see Appendix I, Part II.c.)

Part 4. Final consonant. (For the symbols used, see Appendix I, Part II.d.)

Note: Zero initial and zero final are both coded as ∅.

The contrast between the colloquial and the reading pronunciation of a logograph is shown in the following manner: whenever there is a contrast between these two types of readings, the colloquial pronunciation is coded within parentheses. For example, the logograph 見 in the Xià-mén (Amoy) dialect has these two values: (a) literary kiàn and (b) colloquial kì̃. In our code, these two syllables will be symbolized as follows (□ represents a single space):

C.17 □ P.13P.2KP.3IAP.4N □ (P.13P.2KP.3IZP.4∅)

If there are two or more colloquial variants, but no literary variants, the colloquial readings are not enclosed in parentheses, but are merely separated by one space.

The coding of syllabic nasals presents something of a conceptual problem. For example, the logograph 吳 has the phonetic value ŋ̍ (syllabic velar nasal) in the dialects of Sū-zhōu, Wēn-zhōu, Méi-xiàn, and Guǎng-zhōu. In this case, what is to be considered the initial, the vowel complex, and the final? For this syllable we code zero initial, zero final, and ŋ as the vocalic element. As far as our code is concerned, this treatment accords with the general pattern: only nonvocalic segments can appear as initials or finals. However, if we are interested in finding the reflexes of the Ancient Chinese initial 疑 (usually reconstructed as *ŋ), then the nature of the vocalic segment becomes relevant. Our program must, therefore, instruct the computer not only to examine Part 2 of each cell (in which the initial is coded), but also the contents of Cell 3 in each dialect which has syllabic nasals in its inventory.

2. General notes on the coding format.

In our code each column of the Zìhuì begins with an asterisk and ends in a slash. The asterisk is always punched in column 1 of an IBM card. A cell is specified by C.X, where X stands for the number of the cell. For example, cell one is coded C.1, cell seven is C.7. The cell address, C.X, is always preceded and followed by at least one blank space, except for C.1, which is immediately preceded by an asterisk. Only the first seventy-two columns of an IBM card carry the codes; the last eight columns are reserved for the card identification number. A part (i.e. a division within each cell) is coded P.X, where X stands for the number of the part. Thus, part one is coded P.1, part two is coded P.2, and so on. For example, the coding of the column for the logograph (p. 64) will look like this: *C.1 ☐ 4822A ☐ C.2 ☐ P.1NMP.2KAIP.33DP.44P.54P.6K ☐ C.3 ☐ P.12P.2TCP.3IP.40 ☐ C.4 ☐ P.1, etc. ... ☐ C.9 ☐ P.14P.2TCP.3IE3P.Q ☐ (P.12P.2KP.3E3IP.40) ☐/

☐ represents a significant space in the code. As stated earlier, whenever our code specifies a space between symbols, we have left at least two spaces blank. This extra spacing allows us to make corrections on the IBM cards without shifting the data to preserve the spacing required by the code format. The symbols which appear in Cells 2-19 are listed in Appendix I.

3. The computer program[7]

3.0. Purpose. The Chinese Dictionary Program is designed to read a Chinese dictionary, Hànyǔ fāngyīn zìhuì, from cards and then print tables consisting of certain parts of certain entries. Parts of entries are printed only if the whole entry fulfills specified requirements. At the

present time, specified parts must contain one of several possible choices. The program is organized so that the matching and printing specifications may be easily and clearly given. The program can also make a thorough check of the cards to make sure they are punched properly.

3.1. Organization. The entire Chinese Dictionary Program is actually a collection of small programs built around the main dictionary-search program. This main dictionary program reads in the Hànyǔ fāngyīn zìhuì and prints specified parts of an entry if specified conditions are met. The subsidiary programs do such things as reading in the print specifications and converting them into a form usable by the main dictionary routine, or reading in the match specifications and converting them for use by the main dictionary routine. These separate programs are all tied together by a master program called the Executive. The Executive Program reads control cards and executes programs depending on what the control card says. The standard form of control cards has a '$' in the first column followed by a name. For example:

$DICT. This calls the main dictionary program. The basic operation of this program has already been explained. The cards of the Hànyǔ fāngyīn zìhuì must immediately follow. The program returns control to the Executive Program when a $-card is read.

$PRINT. This sets up the Print Table. All, or any number of, parts of a cell may be specified on a single specification card. The total number of parts is limited only by how much room is on the printer line (see $WIDTH). The following are some examples:

C.1	(print all of Cell 1)
C.2_P.5__SUBRHYME	(print Cell 2, Part 5)

C.12_P.2_P.3__INITIAL_AND_NUCLEUS (print Parts 2 & 3 of Cell 12)

C.17 (print all of Cell 17)

The specifications must be in the same order as in the dictionary entry. Exactly one space must appear between each part of the specification. If two consecutive blanks are found, the rest of the card is treated as a comment. The specifications must start in the first column (i.e. a 'C' must appear in col. 1). A limited check is made to see if these rules are followed.

The Print Table routine reads specification cards and builds up the Print Table until a card with a '$' in the first column is found, when it returns to the Executive Program.

$TITLE. This allows a title to be printed. The dictionary program doesn't print a title, but this option may be used. The paper is normally shifted to a new page; then a line is printed. This line is formed by taking together all eighty columns of the first card following and the first forty columns of the second card following. Thus, each $TITLE card may skip to a new page, then print a single line.

If, however, the control card is of the form $TITLE SUPPRESS, then the next $TITLE card will not skip to a new page. This allows two or more lines to be printed at the beginning of a page.

Each line is followed by a single blank line. After the two cards following the $TITLE card are taken care of, the Executive Program takes over. The next $TITLE card after a simple $TITLE card will be printed on a new page.

$WIDTH n. The normal field width for the printout of the main dictionary program is ten characters. This allows twelve parts to

be printed out. The $WIDTH n control card changes this field width to n characters (n is a single digit). Thus, if n is six, twenty parts can be printed out. The value of n controls only the printing, not the comparisons used to determine whether something will be printed.

If n isn't a digit (e.g. a letter), a message will be typed and the field width set to ten.

$PAUSE. This simply halts the execution until the START button is pushed. It gives the operator time to change the input.

$TYPEWRITER INPUT. This causes all subsequent input to be from the typewriter.

$CARD INPUT. This causes all subsequent input to be from the card reader.

$SEQUENCE CHECK. This causes the cards that form the Hǎnyǔ fāngyīn zìhuì to be sequence-checked. If two cards are out of order, a message is typed and the computer halts.

$NO SEQUENCE CHECK. This stops sequence checking.

$LIST CONTROL CARDS. This causes all subsequent control cards to be typed on the typewriter.

$UNLIST CONTROL CARDS. This stops the listing of control cards.

$END. This signals the end of the run. A message is typed and the IBM Monitor Program takes over.

3.2. Method. The main dictionary-routine works basically on two tables. The Print Table tells what specific parts of each entry are to be printed in the case of a successful match. The Match Table tells what things should match. Each entry in the Match Table consists of a part

number (e.g. $\overline{0}25$ stands for Cell 2, Part 5) and a number of choices. If the specified part of the entry exactly matches any of the choices in the corresponding entry in the Match Table, then the dictionary entry remains as a possible match. If the specified part of the dictionary fails to match any of the choices, the dictionary entry is rejected, and a new entry is considered.

The main dictionary-routine may also be changed so that it prints or types a message if the dictionary entry fails to match. This is particularly useful when the Match Table specifies all legal possibilities, so that any incorrect cards may be found. Though the dictionary routine normally does a partial check on the entries, it cannot catch all of the mistakes without this change. The change may be easily made by means of a control card.

If a mistake is found, the typewriter types a message, e.g. ERROR 02, and stops. When the START button on the console is pushed, the typewriter will type out the offending card and halt again. If the START button is pressed again, the computer will ignore the error and continue as if it were correct. A table of possible errors, their meanings, and possible treatments follows:

Error Table

Number	Routine	Explanation
01	DICT	Neither a 'C' or a 'P' precedes a '.'
02	DICT	A ' ' doesn't follow a 'C.nn'
03	DICT	A '.' doesn't follow a 'C.nP' or a 'C.nnP'
04	DICT	Sequencing error
05	PRINT	First character isn't a '$' or a 'C'
06	PRINT	A '.' doesn't follow a 'C'
07	PRINT	A ' ' doesn't follow a 'C.nn'
08	PRINT	A '.' doesn't follow a 'C.n P' or a 'C.nn P'
09	PRINT	A ' ' doesn't follow a 'C.nn P.n' or a 'C.n P.n'
10	MATCH	First character isn't a '$' or a 'C'
11	MATCH	A '.' doesn't follow a 'C'
12	MATCH	A ' ' doesn't follow a C.nn'
13	MATCH	A 'P' doesn't follow a 'C.n' or a 'C.nn'
14	MATCH	A '.' doesn't follow a 'C.n P' or a 'C.nn P'
15	MATCH	A character that isn't a ' ' or a ',' is between choice fields
16	MATCH	The end of the card was reached while forming a choice
17	MATCH	A choice field was too long

Other possible errors:

(1)	C.1__THE LOGOGRAPH	(incorrect order)
	C.4	
	C.2__ANCIENT_CHINESE_CLASSIFICATION	
(2)	C.3__P.3_P.4__NUCLEUS_AND_ENDING	(P.3_P.4 are treated as a comment)
(3)	C.2_P.1_P.4_P.6_P.5	(incorrect order)
(4)	C.1Ø	(letter O instead of digit 0)
(5)	C.2_P.I_P.5____PEKING_DIALECT	(letter I instead of digit 1) (too high a part number)

Setting up the Match Table

The Match Table tells the program what the parts of a dictionary entry should be. If all parts of an entry match the specifications, the dictionary routine will print out the entry (as specified by the Print Table). There may be a unique specification for a particular part of a cell, or there may be several choices. If a part of a dictionary routine matches any of the choices, it remains a possible match, and will be printed when the end of the entry is reached. If a part of an entry fails to match any of the choices, the entry is skipped and a new entry is begun.

The routine to set up the Match Table is called by a $MATCH card.

This card must be followed by cards specifying what parts must match with which choices. Each card gives information about one part. For example:

C.1_____/1572/,___/3632/,___/6011/,__/5006/

C.2_P.1___/WO/,_____/XG/

C.2_P.2___/KAI/

The specifications must be in the same order as in the dictionary entry. Exactly one space must occur between the C-part and the P-part of the specification. The specifications must start in the first column (i.e. a 'C' must appear in col. 1).

When the Match Table routine reads a card with a '$' in the first column, it returns to the Executive Program.

4. Errors in the Zìhuì.

As Shī Wén-táo pointed out in his review (1963), the Zìhuì contains numerous errors of various types. In the course of our coding, we took into account all the errors noticed by Shī and have corrected those for which he gives the necessary correction. Unfortunately, Shī did not have the opportunity to weed out all the errors in the Zìhuì, but merely listed the types of errors which occur, giving several examples for each. Therefore, we had to be on the alert to catch the remaining errors and correct them to the best of our ability. A list of the errata in the Zìhuì is given in Appendix III of this report. It includes the specific errors mentioned by Shī, as well as those discovered by our coders.

Some of the errors were easily corrected. Others, although we noted the error involved, were not corrected, since to do so would require a consultation with native speakers of the pertinent dialects. In many cases, our errata list merely states that a certain form may be erroneously

transcribed or that some element is missing, but we were unable to provide the correct form. In those cases where an element is omitted, we have coded the missing part as XX; thus, the computer will be able to retrieve all incomplete entries, and we shall be able to correct them in the future by consulting Chinese informants. We shall simply have to go through our errata list and attempt to correct other types of errors as the occasion arises.

Most of the errors found in the Zìhuì can be classified under the following headings:

(1) Omission.

Tone marks are omitted quite frequently. There is reason to suspect that in many cases a nasalization mark is missing, but there was no way of checking this supposition.

(2) Miswritten or wrong logographs in Cell 2.

There are numerous cases of this throughout the Zìhuì. For example, on page 29, Cell 2, Part 6 for the logograph 若 has 曰 instead of 日 ; on page 25, Cell 2, Part 6 for the logograph 挪 has 沼 instead of 泥 .

(3) Mismatch between the rimeme (攝) and the subrime (韻).

For example, on page 38, Cell 2, Parts 1 and 5 of the logograph 掘 do not match, since 月 is a subrime of 山 攝 , not of 臻 攝 .

(4) Mismatch between a dialect form and the phonetic inventory of the same dialect (as listed in the introductory chapter of the Zìhuì).

This type of mismatch often involves tones as well as the segmental elements. For example, on page 4, the logograph 拿 in the Tài-yuán dialect (Cell 6) is listed as being 陽平 in tone, whereas, according to the inventory of tones for this dialect, there is no distinction between 陽平 and 陰平 . On page 10, the logograph 圧 is listed as having the

final -iəp in Cháo-zhōu, but the inventory of finals does not list this final for Cháo-zhōu. According to Shī Wén-táo, in some instances the phonetic inventory is incomplete as it stands and the dialect forms recorded in the Zìhuì are actually correct.

(5) Haphazard handling of the doublets in Ancient Chinese.

It appears that the editors of the Zìhuì either ignored or did not know how to deal with the problem of doublets, i.e. those logographs which had more than one phonetic value in Ancient Chinese and were accordingly listed under different categories in the rime books. In most cases, the editors give in Cell 2 that category which most of the dialect forms fit, without noting that some of the dialects have a form which reflects another Ancient Chinese value of the logograph in question. For example, on page 159, the logograph 灸 is listed as having 上 as its tone in Ancient Chinese and belonging to the subrime 有 . However, in many dialects this word is read as 去聲 : Hàn-kǒu, Wēn-zhōu, Cháng-shā, Nán-chāng, Guǎng-dōng and Cháo-zhōu. If we consult the Guǎng yùn we discover that 灸 is listed there under the subrime 有 (上聲), as well as the subrime 宥 (去聲). Thus, for the purposes of establishing sound correspondences it would be better to split such matrices into two or more matrices, each reflecting a different Ancient Chinese classification.

This problem is further complicated for those logographs which have two different categorizations in the Ancient Chinese rime books: the editors of the Zìhuì have split these into two separate matrices without regard to the behavior of these doublets in other Chinese dialects. This overemphasis on the written representation of Chinese morphemes as well as the overall orientation towards Peking Mandarin results in a skewed

picture of the phonological correspondences among individual dialects: in many dialects the phonological distinction between the doublets is erased, perhaps through analogy; the statistically dominant phonological value of the logograph may replace the less-used phonological value of the same logograph.

For example, in Peking Mandarin the logograph 教 has two pronunciations: $_{c}$tɕiau (平聲) in the expression 教書 'to teach', and tɕiau^{3} (去聲) in the expression 教育 'education'. This tonal distinction is duly reflected in the Guǎng yùn, which classifies the first instance as belonging to the subrime 肴 (平聲) and the second instance as belonging to the subrime 效 (去聲). However, a glance at the two matrices for this logograph in the Zìhuì (pages 144 and 145) shows that although 教 in the expression 教育 is reflected as a 去聲 in every dialect listed, in the case of 教 in 教書 , the Mandarin dialects, the Wēn-zhōu dialect, and the Méi-xiàn dialect reflect the 平聲 , but none of the other dialects reflect the expected tone, namely 去聲 . Obviously, in the latter dialects the two values of this logograph merged into one which is pronounced 去聲 .

From the point of view of comparative morphology, this and similar phenomena are, to be sure, highly interesting. On the other hand, if we are primarily interested in tabulating phonological correspondences, the arrangement of the data in the Zìhuì is very unsatisfactory, and, in the case described above, would cause the computer to tabulate the following tone correspondence which is obviously invalid: Ancient Chinese 平聲 : Mandarin dialects 平聲 : Méi-xiàn dialect 平聲 : Wēn-zhōu dialect 平聲 : other dialects 去聲 .

Of course, the number of logographs showing such faculty correspondences will probably be so small that it will be easy to spot these anomalies. For our purposes, however, it would have been much more desirable to leave blank those cells which show morphological levelling rather than the regular phonological development.

The problem of the doublets in the Zìhuì is so complex that it was not possible to solve it during the course of the coding. One way out of this dilemma is to leave the matrices as they are, and then by utilizing a special computer program, to extract and tabulate all the data on the doublets and the irregular correspondences which may involve doublets. Once such a tabulation is available, the necessary splitting of the matrices and the deletion of the irregular forms can be effected.

(6) Faulty cognates.

There are cases where one would suspect that a dialect form does not belong in a particular matrix since it shows very unusual sound correspondences. For example, on page 41 we have the logograph 子 which in most of the dialects has ts as initial (and according to Guǎng yùn, has the initial 精 in Ancient Chinese, usually reconstructed as *ts); one of the colloquial forms in the Amoy dialect, however, is ^{c}kiã̃ (along with ^{c}tsi and ^{c}dzi), which does not seem to belong in the same phonological category. Again, on page 70 the Amoy colloquial reading of the logograph 一 is given as tsit$_{5}$, but no other dialect shows ts as an initial, and the Guǎng yùn classification indicates that the Ancient Chinese initial was a glottal stop. This Amoy form may be a reflex of some other Chinese word cognate to Tibetan chig ('one') or a loan word from some Tibeto-Burman language. (Judging from several Amoy forms, I suspect that the finals -Ik and -it are sometimes confused with each other.)

Nevertheless, it seemed best not to alter or discard such examples. They can be singled out for special study after the computer tabulates all apparent irregularities.

In addition to the above-mentioned types of errors, there are numerous minor errors of omission and commission. The IPA symbols are often written cursively; what is more confusing, the same IPA symbol varies radically in its graphic representation. Thus, i varies from $\dot{\imath}$ to $\imath$, $\dot{\imath}$ to i (see pages 99, 102, and 123 of the Zìhuì).

The general inadequacy of the Zìhuì, however, lies in its basic orientation to the Chinese logograph. Some of the errors which are caused by this orientation were already discussed in connection with the problem of doublets; another type of error will illustrate this point further.

On page 166, the logograph 担 is listed as having the following Ancient Chinese classification: 咸开一上旱端 . However, 旱 is a subrime of the 山 攝 , not of 咸 攝 . Thus, there is a mismatch between the rimeme (攝) and the subrime (韻員). It turns out that Guǎng yùn lists two different logographs, 担 and 擔 ; the former is classified as 山开一上旱端, the latter as 咸开一去闞端. Now, 担 is currently used as an abbreviated form of 擔 , since in Peking Mandarin the pronunciation of the two logographs is identical, although in Ancient Chinese the two words ended in -n and -m, respectively. It appears that in the Zìhuì 担 is the abbreviated form of 擔 , since those dialects which have preserved the distinction between -n and -m show -m as the final.

This raises the question: are there many more mistakes which have been caused by the use of abbreviated logographs? We must keep in mind that the abbreviations were usually devised on the basis of Mandarin pronunciation, whereas many dialects preserve more original distinctions than

Mandarin. Second, if written logographs were used to elicit phonetic data from dialect speakers, how much care was taken to see that the informants did not merely give the pronunciation of the phonetic element of a logograph if the morpheme represented by a particular logograph was not currently used in their dialect? In the above case, both 旦 and 詹 occur frequently as phonetic elements in logographs, and confusion is very likely to occur.

On the whole, however, the bulk of the data in the Zìhuì appears to be valid in spite of the methodological faults involved in its compilation, and will undoubtedly yield much relevant information concerning phonological developments in Chinese dialects.

Appendix I.

Computer code for the Hànyǔ fāngyīn zìhuì.

PART I (The code for Cell 2)

(a) The 16 shè (Cell 2, Part 1):

遇 =	NO	假 =	WO
果 =	OO	蟹 =	WI
止 =	NI	效 =	WU
流 =	NU	咸 =	WM
深 =	NM	山 =	WN
通 =	NG	臻 =	XN
宕 =	OG	江 =	WG
曾 =	PG	梗 =	XG

(b) Kāi-kǒu vs. hé-kǒu (Cell 2, Part 2):

開口 = KAI

合口 = HE

(c) The four děng (Cell 2, Part 3):

一等 = 1D

二等 = 2D

三等 = 3D

四等 = 4D

(d) The four tones (Cell 2, Part 4):

平 = 1

上 = 2

去 = 3

入 = 4

(e) The yùn or subrimes (Cell 2, Part 5). (Arranged according to shè):

(1)遇攝 (NO)

模 = 1
姥 = 2
暮 = 3
魚 = 4
語 = 5
御 = 6
虞 = 7
廣 = 8
遇 = 9

(2)果攝 (OO)

歌 = 1
哿 = 2
箇 = 3
戈 = 4
果 = 5
過 = 6

(3)假攝 (WO)

麻 = 1
馬 = 2
禡 = 3

(4) 蟹攝 (WI)

咍 = 1
海 = 2
代 = 3
泰 = 4
皮 = 5
賄 = 6
隊 = 7
皆 = 8
駭 = 9
怪 = 10
佳 = 11
蟹 = 12
卦 = 13
夬 = 14
祭 = 15
廢 = 16
齊 = 17
薺 = 18
霽 = 19

(5) 止攝 (NI)

支 = 1
紙 = 2
寘 = 3
脂 = 4
旨 = 5
至 = 6
之 = 7
止 = 8
志 = 9
微 = 10
尾 = 11
未 = 12

(6) 效攝 (WU)

豪 = 1
皓 = 2
號 = 3
肴 = 4
巧 = 5
效 = 6
宵 = 7
小 = 8
笑 = 9
蕭 = 10
篠 = 11
嘯 = 12

(7) 流攝 (NU)

侯 = 1
厚 = 2
候 = 3
尤 = 4
有 = 5
宥 = 6
幽 = 7
黝幼 = 8
幼 = 9

(8) 咸攝 (WM)

覃 = 1
感 = 2
勘 = 3
合 = 4
談 = 5
敢 = 6
闞 = 7
盍 = 8
咸 = 9
豏 = 10
陷 = 11
洽 = 12
銜 = 13
檻 = 14
鑑 = 15
狎 = 16
豔 = 17
琰 = 18
豔 = 19
葉 = 20
嚴 = 21
儼 = 22
釅 = 23
業 = 24
凡 = 25
范 = 26
梵 = 27
乏 = 28
添 = 29
忝 = 30
栝 = 31
帖 = 32

(9) 深攝 (NM)

侵 = 1
寢 = 2
沁 = 3
緝 = 4

(10) 山攝 (WN)

寒 = 1
旱 = 2
翰 = 3
易 = 4
桓 = 5
緩 = 6
換 = 7
末 = 8
刪 = 9
潸 = 10
諫 = 11
鎐 = 12
山 = 13
產 = 14
襉 = 15
黠 = 16
仙 = 17
獮 = 18
線 = 19
薛 = 20
元 = 21
阮 = 22
願 = 23
月 = 24
先 = 25
銑 = 26
霰 = 27
屑 = 28

(11) 通攝 (NG)

東 = 1
董 = 2
送 = 3
屋 = 4
冬 = 5
腫 = 6
宋 = 7
沃 = 8
燭 = 9
鍾 = 10
用 = 11

(12) 臻攝 (XN)

痕 = 1
很 = 2
恨 = 3
紇 = 4
魂 = 5
混 = 6
恩 = 7
沒 = 8
臻 = 9
櫛 = 10
真 = 11
軫 = 12
震 = 13
諄 = 14
準 = 15
稕 = 16
術 = 17
質 = 18
欣 = 19
隱 = 20
焮 = 21
迄 = 22
文 = 23
吻 = 25
問 = 26
物 = 27

(13) 宕攝 (OG)

唐 = 1
蕩 = 2
宕 = 3
鐸 = 4
陽 = 5
養 = 6
漾 = 7
藥 = 8

(14) 江攝 (WG)

江 = 1
講 = 2
絳 = 3
覺 = 4

(15) 曾攝 (PG)

登 = 1
等 = 2
嶝 = 3
德 = 4
蒸 = 5
極 = 6
證 = 7
職 = 8

(16) 梗攝 (XG)

庚 = 1
梗 = 2
映 = 3
陌 = 4
耕 = 5
耿 = 6
諍 = 7
麥 = 8
清 = 9
靜 = 10
勁 = 11
昔 = 12
青 = 13
迥 = 14
徑 = 15
錫 = 16

(f) The initials or zìmǔ (Cell 2, Part 6):

i. Labials

幫 = P
滂 = P'
並 = B
明 = M

非 = F
敷 = F'
奉 = V
微 = MG

ii. Dentals

端 = T
透 = T'
定 = D
泥 = N

精 = TS
清 = TS'
從 = DZ
心 = S
邪 = Z

iii. Palatals and retroflexes

知 = TJ
徹 = TJ'
澄 = DJ
日 = NJ

章 = TSJ
昌 = TSJ'
船 = DZJ
書 = SJ
禪 = ZJ

莊 = TSH
初 = TSH'
崇 = DZH
生 = SH
俟 = ZH

iv. Velars, etc.

見 = K
溪 = K'
群 = G
疑 = NG

影 = '
曉 = X
匣 = GH
云 = GHJ
以 = ø (Zero)

v. Lateral

來 = L

Part II (The code for Cells 3-19)

(a) Tones (Part 1):

(陰)平 = 1　　(陰)上 = 2　　(陰)去 = 3　　入 = 4

陽平 = 1B　　陽上 = 2B　　陽去 = 3B　　陽入 = 4B

中陰入(Cantonese) = 4C

(b) Initials (Part 2):

Zero initial: ∅ (zero)

Stops and affricates:

Labial	Alveolar	Palatal	Retroflex, etc.	Postvelar
p = P	t = T		tʂ = TSR	k = K
p' = PH	t' = TH		tʂ' = TSRH	k' = KH
pf = PF	ts = TS	tɕ = TC	tʃ = TSP	kw = KW
pf' = PFH	ts' = TSH	tɕ' = TCH	tʃ' = TSPH	kw' = KWH
b = B	d = D			g = G
ɓ = BQ	ɗ = DQ			
	dz = DZ	dʑ = DZR		

Note: Noninitial H always stands for aspiration.

Fricatives:

Labial	Alveolar	Palatal	Retroflex, etc.	Postvelar
Φ = FF	s = S	ɕ = C	ʂ = SR	x = X
β = VV	z = Z	j = J	ʐ = ZR	ɣ = XV
f = F			ʃ = SP	h = H
f^h = FH	s^h = SH	$ɕ^h$ = CH	$ʂ^h$ = SRH	
v = V	z^h = ZH		ʒ = ZZ	ɦ = HV

Liquids:

w = W	l = L	ɹ = R	
	ɬ = LS		
	$l^{ɦ}$ = LH		

Nasals:

m = M	n = N	ɲ = NJ	ŋ = NG
$m^{ɦ}$ = MH	$n^{ɦ}$ = NH	$ɲ^{ɦ}$ = NJH	$ŋ^{ɦ}$ = NGH

(c) Vocalic nuclei (Part 3):

i = I	y = Y	u = U
I = I1	Y = Y1	U = U1
ɪ = I2	ʏ = Y2	o = O
ɿ = I3		ɤ = O1
e = E	ø = O3	ɔ = O2
E = E1		ɯ = U3
ɛ = E2	oe = OE	
ə = E3	ɵ = O4	
ɚ = E4	ɞ = O4 (Fù-zhōu)	
ɜ = E5		
æ = AE		ɑ = A1
a = A		ɒ = A2

Syllabic nasals and syllabic liquid:

m̩ = MM	n̩ = NN	ŋ̩ = NNG	l̩ = LL

Note: A nasalized vowel is coded vowel plus Z. Vowel length (:) is coded vowel plus W.

(d) Endings (Part 4):

m = M	n = N	ŋ = NJ	ŋ = NG	
p = P	t = T	c = KJ	k = K	ʔ = Q

Note:

Zero ending is coded explicitly as ∅ (zero).

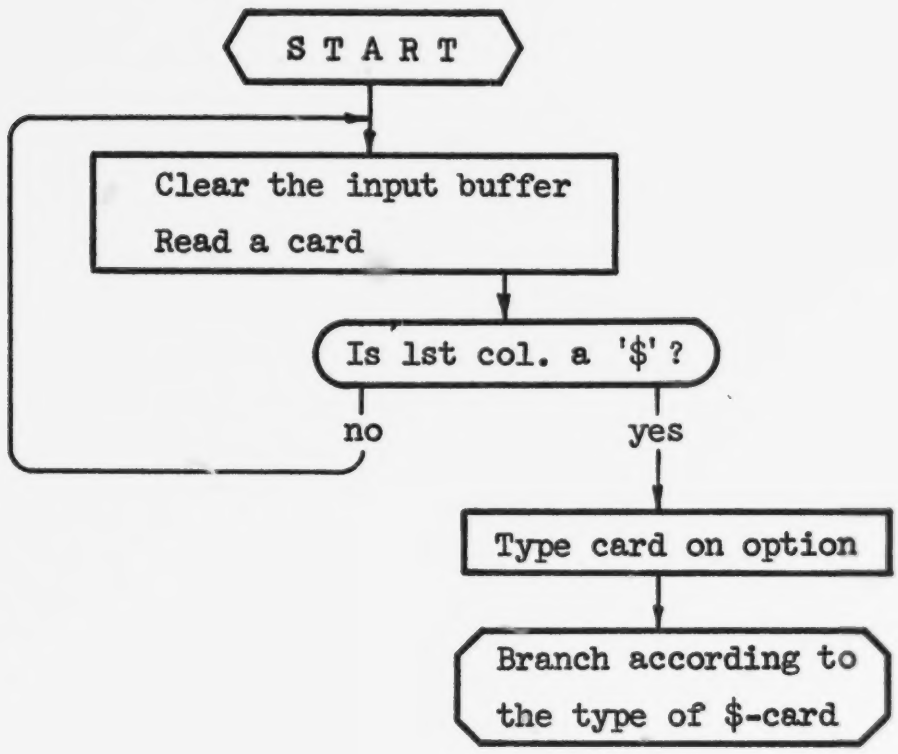
Special symbols:

XX in any cell or part indicates that an element is either missing or is incorrectly specified in the Zìhuì, but that we were unable to supply the missing element or correct the error.

Appendix II.

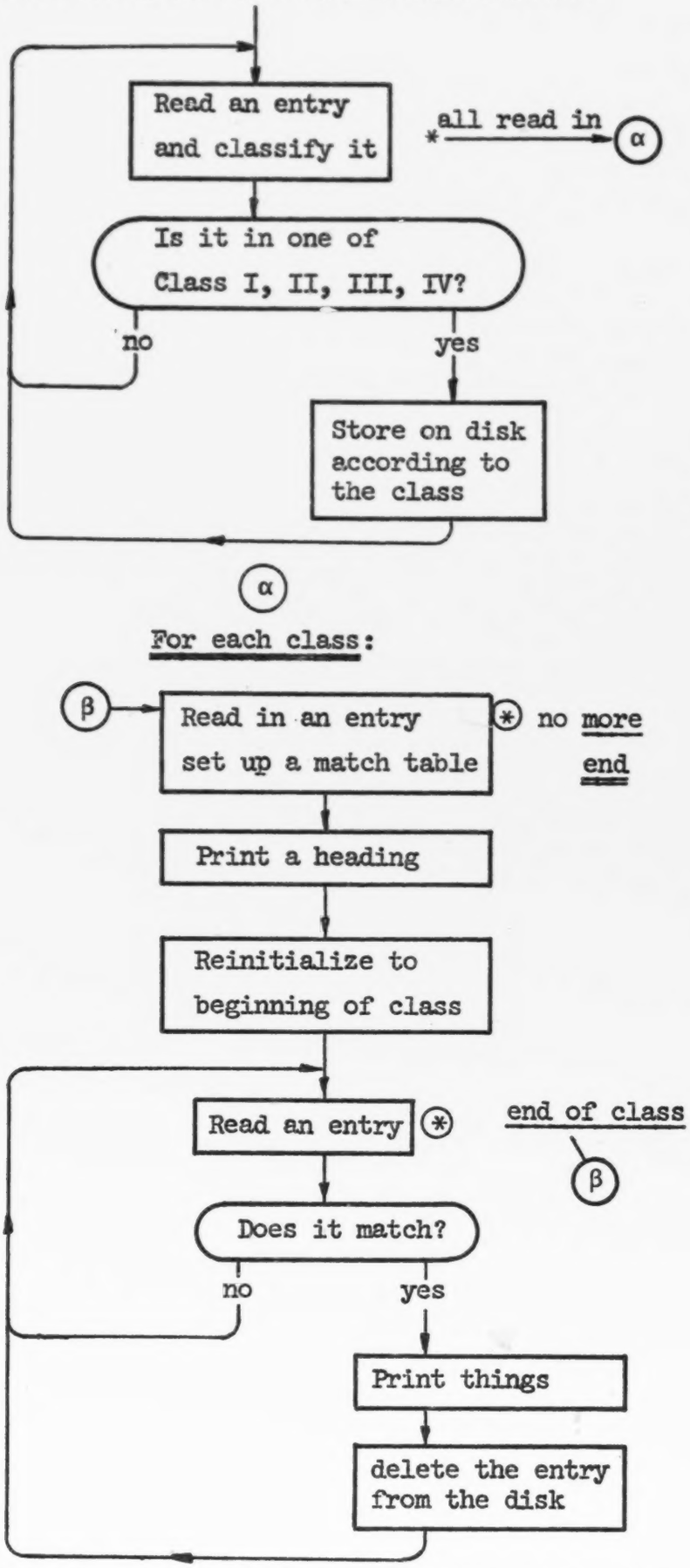
Flowcharts for the computer program

David Forthoffer



$ TITLE	INIT.T
$ PRINT	INIT.P
$ MATCH	INIT.M
$ DICT	DICT
$ TYPEWRITER INPUT	TY.IN
$ CARD INPUT	CRD.IN
$ SEQUENCE CHECK	SEQCH
$ NØ SEQUENCE CHECK	NSEQCH
$ LIST CØNTRØL CARDS	LISTCC
$ UNLIST CØNTRØL CARDS	UNLIST
$ WIDTH	WIDTH
$ PAUSE	PAUSE
$ END	THEEND

SIMPLIFIED FLOW CHART FOR THE MASTER PROGRAM



NOTE: This program is designed to classify any segment of data into categories with an unlimited number of subcategories with matching items.

Appendix III

A list of the errata in the Hànyǔ fāngyīn zìhuì.

Hsin-i Hsieh

page	character	cell	comment
1	拔	潮州	-uek does not appear in the inventory of finals
2	法	潮州	-uap does not appear in the inventory of finals.
4	拿	太原	ˏna should be ˊna.
5	鍘	潮州	-oik does not appear in the inventory of finals.
9	价	福州	lacks tone mark.
10	夏韻		上 should be 去 . Guǎng yùn gives 胡駕切.
10	圧	潮州	-iəp does not appear in the inventory of finals.
11	爪	濟南	-ɒ does not appear in the inventory of finals.
11	爪	潮州	z- does not appear in the inventory of initials.
11	爪	潮州	-iəu does not appear in the inventory of finals.
14	捽	雙峯	lacks tone mark.
15	哲	潮州	-iek does not appear in the inventory of finals.
19	克	南昌	lacks tone mark.

page	character	cell	comment
19	刻	南昌	lacks tone mark.
19	刻	西安	lacks tone mark.
19	客	南昌	lacks tone mark.
24	夺	福州	-uəʔ does not appear in the inventory of finals.
26	挪	2	沼 should be 泥 .
26	糯	2	沼 should be 泥 .
26	鑿	潮州	-ɔʔ does not appear in the inventory of finals.
27	左	2	迦 should be 哿 .
27	梭	西安	lacks tone mark.
28	着	潮州	-ieʔ does not appear in the inventory of finals.
29	濁	2	幵三 should be 幵二 ; 藥 should be 覺 .
29	若	2	曰 should be 日 .
29	弱	2	曰 should be 日 .
29	國	西安	-ue does not appear in the inventory of finals.
31	和	濟南	ɕxuəˀ should be ɕxuə and xuəˀ. ɕxuə is from the reading 音禾 and xuəˀ from the reading 胡卧切 in Guǎng yùn.
32	爹	福州	-ie does not appear in the inventory of finals.
33	皆	濟南	-iɤ does not appear in the inventory of finals.

page	character	cell	comment
33	阶	濟南	-iɤ does not appear in the inventory of finals.
34	接	潮州	-iep does not appear in the inventory of finals.
35	且	濟南	ᶜtɕ'ie could be a mistake for ᶜtɕ'ie .
38	掘	2	臻 should be 山 .
44	脂	厦門	lacks tone mark.
44	只一只	雙峯	lacks tone mark.
44	执	潮州	-ip does not appear in the inventory of finals.
46	吃	北京	lacks tone mark.
46	齒	濟南	ts' should be tʂ' .
49	識	北京	a more common pronunciation is ʂɿˀ .
55	米	2	平 should be 上 .
56	堤	2	堤 and 隄 represent the same word, 'bank'. In Guǎng yùn, 隄 has two pronunciations, one with the voiced initial 定 , the other with the voiceless initial 端 . Some dialects reflect the former pronunciation; others reflect the latter. Probably this is why modern reflexes from Ancient 堤 (and 隄) appear somewhat irregular.
59	你	2	平 should be 上 .
59	膩	2	尼 should be 至 .
64	給	濟南	ᶜtɕi should either be ᶜtɕi or ᶜtɕi .
64	剤	福州	-ə does not appear in the inventory of finals.

page	character	cell	comment
65	繼	福州	-iɛ does not appear in the inventory of finals.
65	寄	福州	-iɛ does not appear in the inventory of finals.
67	啓	2	养 should be 聲 .
74	僕	2	遇 should be 通 .
77	斧	福州	-uo does not appear in the inventory of finals.
87	出	濟南	ˌtʂ'uˏ should be ˌtʂ'u .
89	树	2	过 should be 御 .
97	句	2	通 should be 遇 .
101	与	成都	lacks tone mark.
113	北	福州	-œy˧ does not appear in the inventory of finals.
123	惠	2	曾 should be 霽 .
124	桅	2	支 should be 灰 .
127	薄	2	岩 should be 宕 .
128	豹	濟南	po˥ should be pɔ˥ .
131	套	濟南	t'o˥ should be t'ɔ˥ .
134	糙	北京	米造 is usually pronounced ts'au˥ .
139	熬	2	影 should be 疑 .
142	刀	太原	lacks tone mark.
154	瘦	西安	-əu does not appear in the inventory of finals.
158	留	太原	lacks tone mark.
158	究	北京	究 is also pronounced tɕiou˥ .

page	character	c	comment
158	究	藕州	lacks tone mark.
159	灸	2	according to Guǎng yùn, 灸 has both a shǎng shēng and a qù shēng reading.
159	救	藕州	lacks tone mark.
159	舅	藕州	lacks tone mark.
160	囚	太原	$_{c}$tɕ'iou should be $_{c}$tɕ'iou.
160	修	成都	lacks tone mark.
162	右	藕州	lacks tone mark.
162	祐	藕州	lacks tone mark.
162	柚	藕州	lacks tone mark.
162	幼	藕州	lacks tone mark.
162	玤	揚州	lacks tone mark.
162	板	2	潛 should be 潸.
163	拌	福州	-uaŋ does not appear in the inventory of finals.
163	盼	2	澗 should be 襉.
164	判	太原	p'æ˞ should be p'æ̃˞ (?)
164	滿	2	援 should be 緩.
164	帆	潮州	-uam does not appear in the inventory of finals.
164	帆	2	帆 has both píng shēng and qù shēng readings.
166	担	2	based on the information given in Guǎng yùn, 担 is to be marked as 山开一上旱端, and 擔 is to be marked as 咸开一去闞端. It seems that in the Zìhuì 担 is intended to be the abbreviated form of 擔.

page	character	cell	comment
167	探	潮州	-am˥ does not appear in the inventory of finals.
169	攬	揚州	ᶜlɛ should be ᶜlɛ̃ .
169	黯	潮州	-iəm does not appear in the inventory of finals.
170	粘	2	Guǎng yùn also gives 女廉切. The pronunciations of 粘 in 揚州 , 梅縣 , and 潮州 probably reflect this form.
171	棧	2	上 should 去 . However, Guǎng yùn gives three different pronunciations, two in 上 , and one in 去 .
171	展	潮州	-ieŋ does not appear in the inventory of finals.
172	鏟	2	琰 should be 產 .
172	產	2	Both Jí yùn and Guǎng yùn suggest the initial 生 for the character 產 .
178	顛	2	定 should be 先 .
179	典	2	銳 should be 銑 .
183	檢	成都	lacks tone mark.
183	儉	成都	lacks tone mark.
184	漸	2	去 should be 上 .
185	籤	西安	ᶜtɕʻiãˀ should be ᶜtɕʻiã .
185	謙	西安	ᶜtɕʻiãˀ should be ᶜtɕʻiã .
185	迁	西安	ᶜtɕʻiãˀ should be ᶜtɕʻiã .
187	掀	福州	h- does not appear in the inventory of initials.

page	character	cell	comment
189	酽	西安	$_{c}$ia could be a mistake for $_{c}$iã.
190	嚴	西安	$_{c}$ia could be a mistake for $_{c}$iã.
190	研	太原	lacks tone mark.
193	鉆占	2	should read: 山合一去換精.
194	串	2	線 should be 諫.
195	拴	2	should read: 山合四平仙清. Guǎng yùn gives 此緣切 for this character.
195	管	揚州	nasalization mark omitted (?).
195	貫	揚州	nasalization mark omitted (?).
196	寬	太原	nasalization mark omitted (?).
199	圈	2	上 should be 去.
200	选	2	三合三 should be 山合三.
202	噴	2	Guǎng yùn gives both pīng shēng and shǎng shēng for this character.
203	門	成都	-en does not appear in the inventory of finals.
203	悶	成都	-en does not appear in the inventory of finals.
203	墳	揚州	-an does not appear in the inventory of finals.
205	鎮	2	振 should be 震.
209	殯	2	平 should be 去.
210	民	2	直 should be 真.
211	斤	2	殷 should be 欣.
212	尽	揚州	lacks tone mark.

page	character	cell	comment
212	近	揚州	lacks tone mark.
212	勁	揚州	lacks tone mark.
212	勁	潮州	-ẽŋ does not appear in the inventory of finals.
212	侵	廣州	ts'- does not appear in the inventory of initials.
213	勤	2	殷 should be 欣.
213	欣	2	殷 should be 欣.
214	殷	2	殷 should be 欣.
214	隱	2	引 should be 以.
214	隱	揚州	lacks tone mark.
214	引	揚州	lacks tone mark.
216	崙	2	定 should be 來.
217	春	2	章 should be 昌.
226	張	潮州	-ɐ̃ does not appear in the inventory of finals.
227	帳	2	岩 should be 宕.
231	粮	潮州	-ɐ̃ does not appear in the inventory of finals.
232	疆	蕪州	lacks tone mark.
232	薑	蕪州	lacks tone mark.
240	望	潮州	-õ does not appear in the inventory of finals.
240	崩	成都	lacks tone mark.
255	京	2	庚开三 should be 梗开三.

page	character	cell	comment
255	荊	2	庚开三 should be 梗开三.
256	驚	2	庚开三 should be 梗开三.
258	請	2	庚开三 should be 梗开三.
260	应	2	Guǎng yùn gives pīng shēng and qù shēng for the character 应.
260	纓	厦門	iã lacks tone mark.
262	筒	2	走 should be 定.
265	聪	雙峯	ts'ən lacks tone mark.
265	从	西安	-uŋ does not appear in the inventory of finals.
265	松	2	疑 should be 鍾.

References

Dǒng, Tóng-hé. 1954. History of Chinese phonology. Taipei. China Culture Publishing Foundation. No. 26 of Xiàndài guómín jīběn zhīshì cōngshū. (In Chinese.)

Li, Charles N. 1966. Workpaper 3: A coding system for Chinese characters. University of California, Berkeley. (Unpublished.)

_____. 1967. Report on coding and programming information in the dialect dictionary, Hànyǔ fāngyīn zìhuì. (Unpublished report.)

Peking University. 1962. Hànyǔ fāngyīn zìhuì. [A lexicon of the Chinese dialects]. Peking, Wénzì Gǎigé Chūbǎnshè. (In Chinese.)

Shī, Wén-táo. 1963. Review of A lexicon of the Chinese dialects. Zhōngguó yǔwén 123.176-82. (In Chinese.)

Wang, William S-Y. 1966. Workpaper 1: Rime dictionaries. University of California, Berkeley. (Unpublished.)

_____. 1966. Workpaper 2: Dialect dictionaries. University of California, Berkeley. (Unpublished.)

Footnotes

[1] The preparation of this report as well as the project described therein was supported by National Science Foundation Grant GS1430.

[2] Workpaper 2 (Wang 1966) presented a system of coding the phonetic data in the Zìhuì. The problem of coding the Chinese logographs as well as some Ancient Chinese categorizations was left to be solved by Charles N. Li.

[3] The Hànyǔ fāngyīn zìhuì (Peking University 1962) is one of the most comprehensive Chinese dialect surveys available.

[4] Charles Li (Li 1966) devised an alphanumeric code for Chinese logographs which analyzed the logographs into strokes, each specified by a code letter; the relative position of the stroke within the logograph was indicated by reference to the horizontal and vertical axes.

[5] The Chinese Telegraphic Code can be found in, for example, the Modern Chinese-English technical and general dictionary (McGraw-Hill, 1963), Vol. 1.

[6] The choice of Dǒng's reconstruction (Dǒng 1954) was entirely arbitrary; our coding scheme will in no way prejudice the results since it faithfully preserves the existing distinctions without necessarily specifying what constituted these distinctions.

[7] This section of the report, as well as Appendix II, was written by David Forthoffer, a programmer for the project.